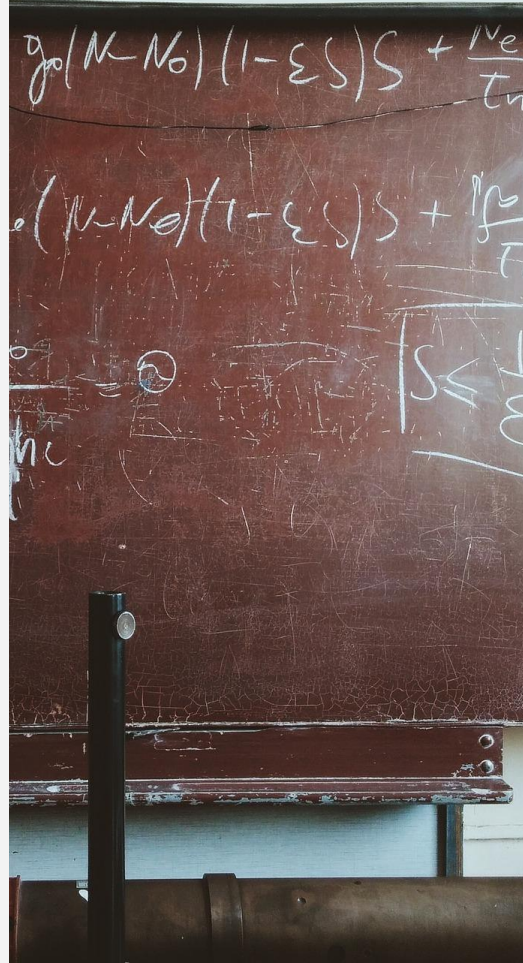


Análise Preditiva I

01

TIPOS DE ANÁLISES



Análise Descritiva

O QUE ACONTECEU?



Analise Diagnóstica

**O QUE ESTÁ
ACONTECENDO?**



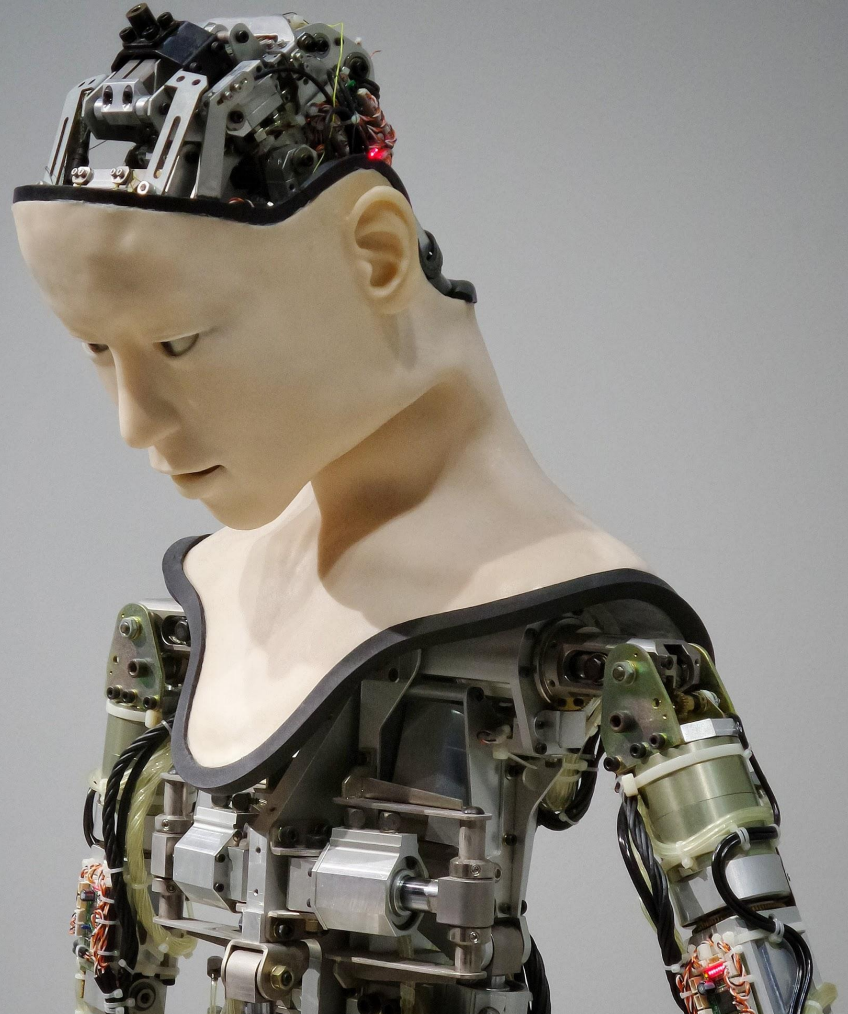
Analise Preditiva

O QUE VAI ACONTECER?

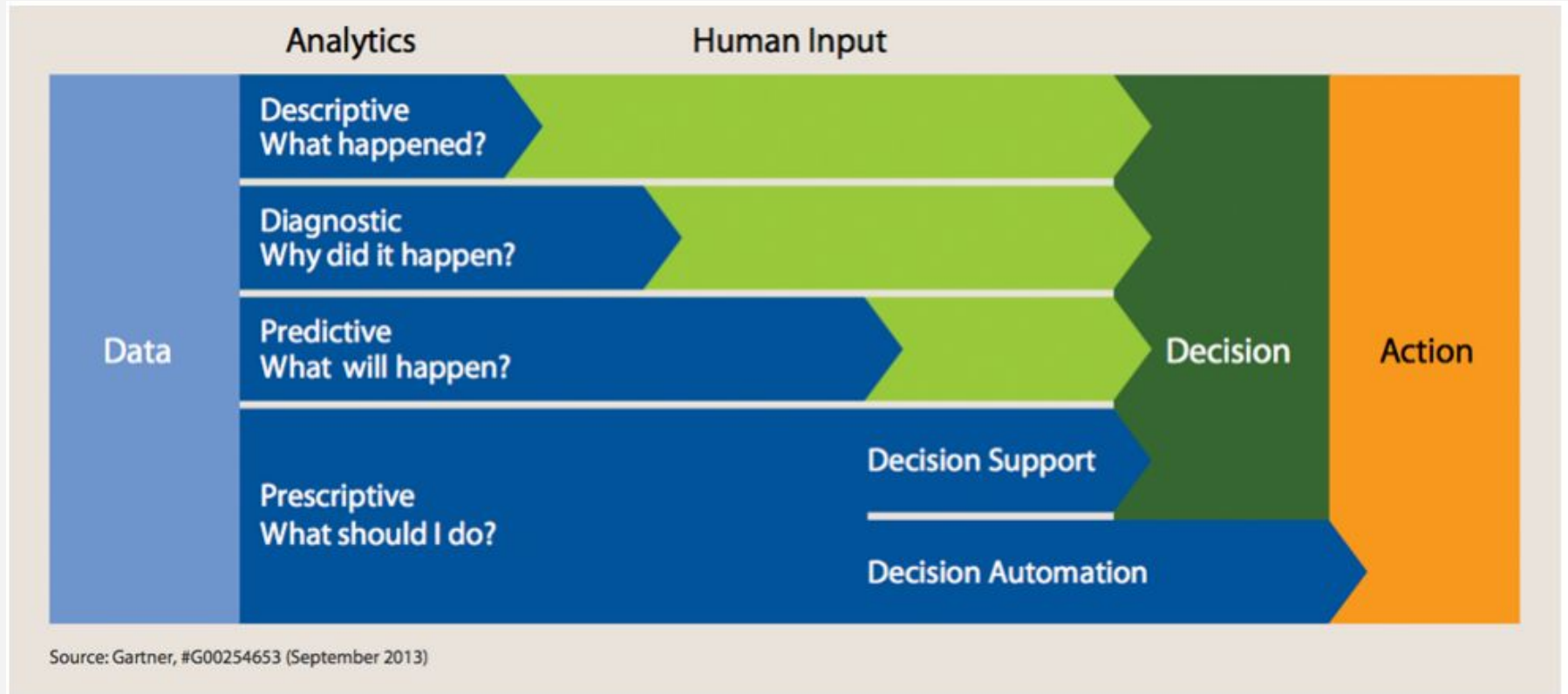


Análise Prescritiva

O QUE DEVEMOS FAZER?

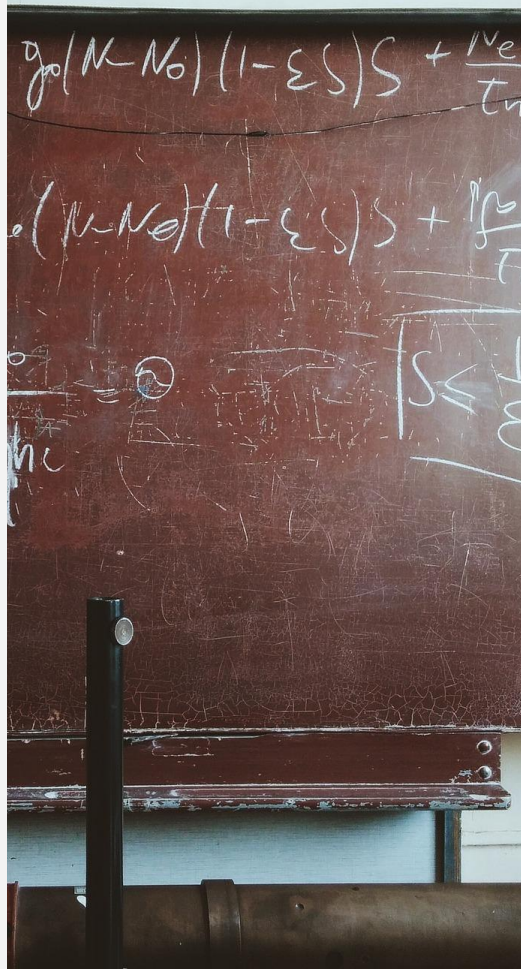


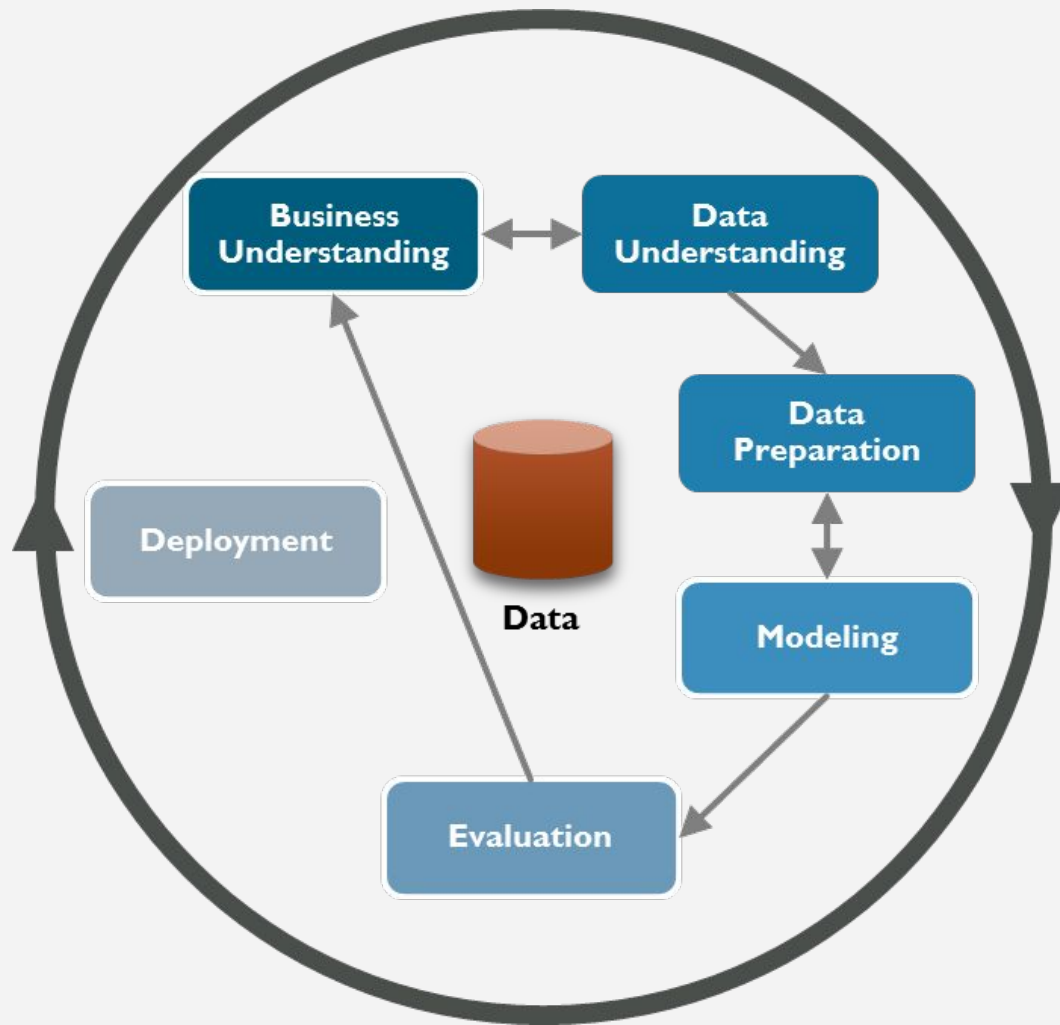
Analises de Dados



02

CRISP





Entendimento de Negócio

- Objetivo do Projeto
- Justificativa
- Matriz de Riscos e Impactos
- Benefícios
- Matriz Custo Benefícios
- Monitoramento (KPIs)

Entendimento dos Dados

- Fontes de Dados
- Custo dos Dados
 - Armazenamento
 - Processamento
- Definição da Variável Alvo
- Definição da Abordagem e Tarefa
 - Supervisionada (Classificação, Regressão, etc)
 - Não Supervisionada (Agrupamento, Associação, etc)

Preparação do Dados

- Tratamento dos Outliers
- Tratamento dos Dados Ausentes
 - Separar Amostra de Treino, Teste e Validação
- Definir método de avaliação

Modelagem

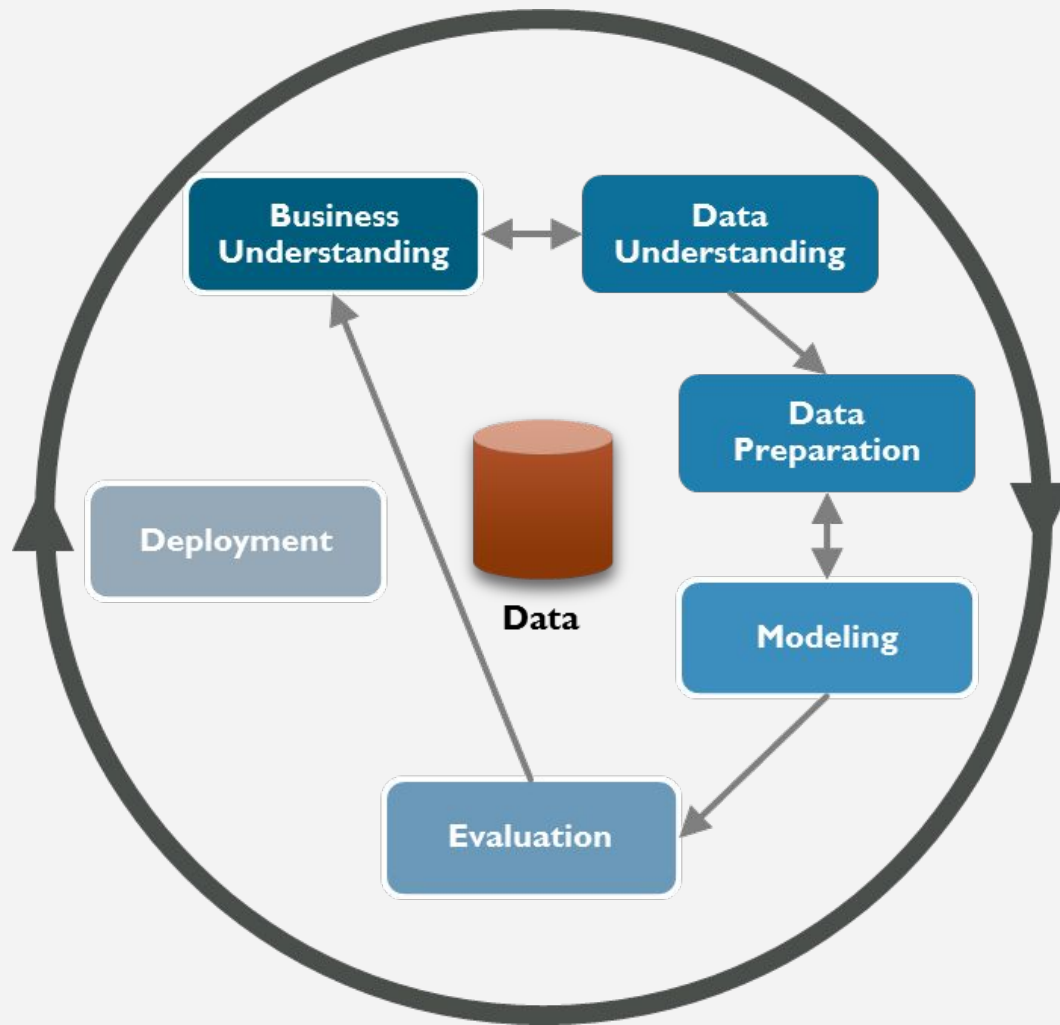
- Listar principais modelos
- Treinar todos os principais modelos

Avaliação

- Avaliação do Modelo
- Avaliação do Custo e Benefício

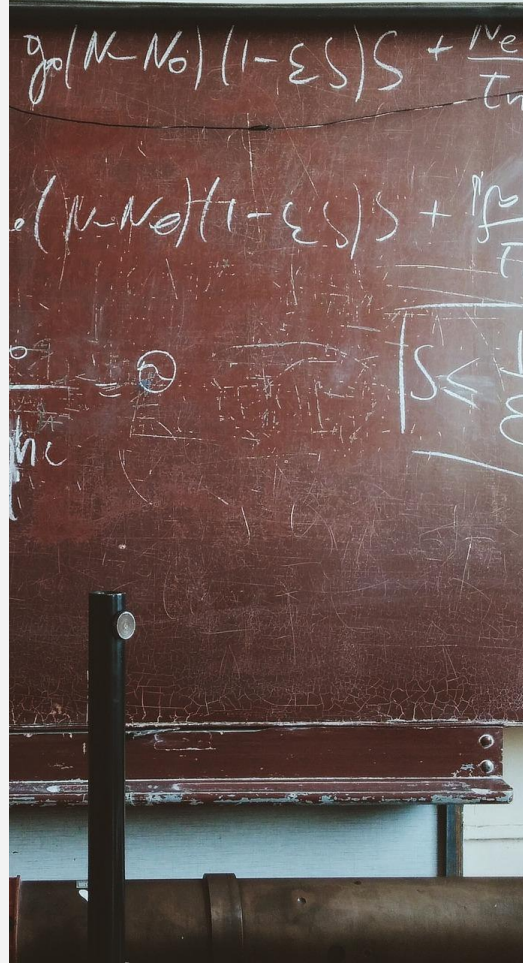
Implantação

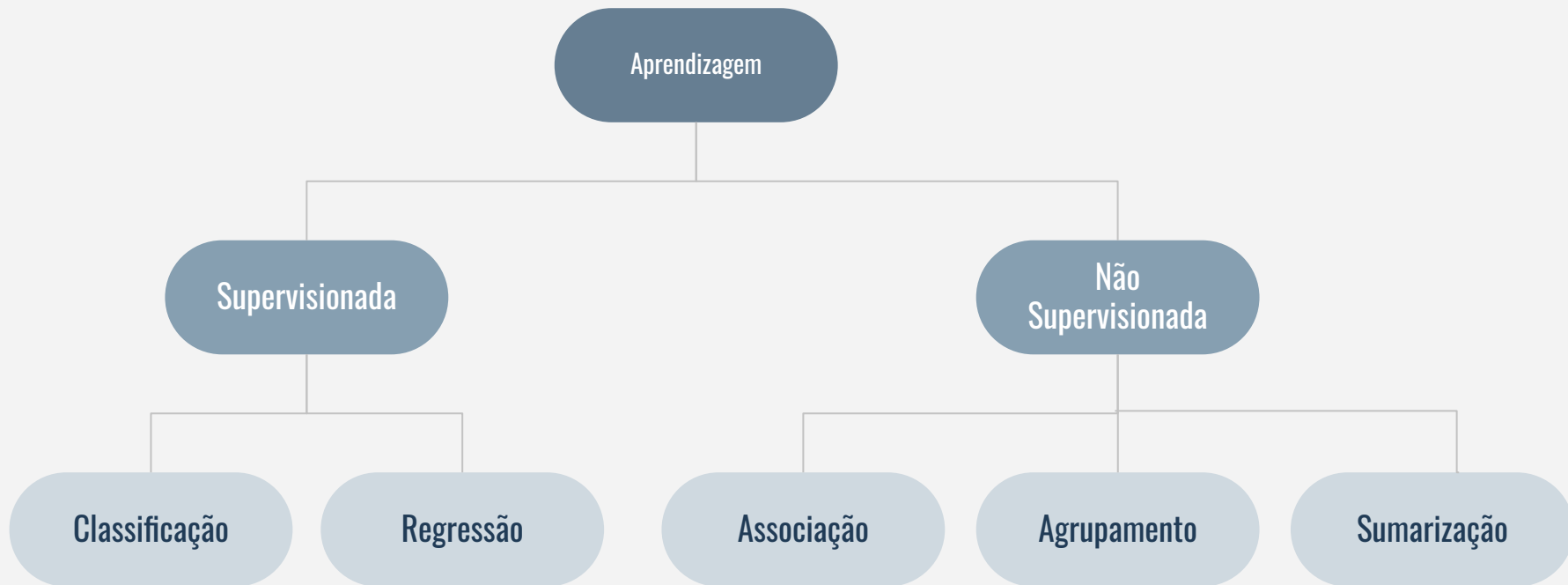
- Plano de Monitoramento
- Plano de Manutenção
- Lições Aprendidas



03

ABORDAGENS PREDITIVAS



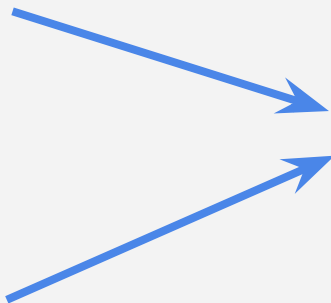




MAÇÃ



BANANA



Supervisionado



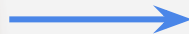
É uma maçã!!!

O QUE É ISSO?





Frutas



Grupo 1



Grupo 2

ABORDAGENS SUPERVISIONADAS



C

Classificação



R

Regressão

Regressão

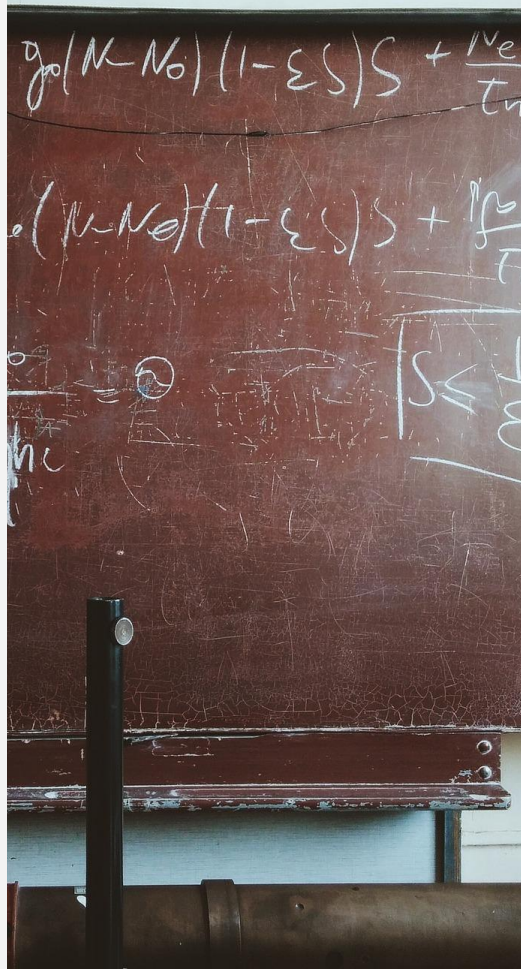
- Qual é o valor da ação amanhã?
- Qual é a máxima de temperatura amanhã

Classificação

- Qual é o fruta é a maçã?
- Qual notícia é de esporte?

03

INTRODUÇÃO A MODELAGEM PREDITIVA (CLASSIFICAÇÃO)



Será que o cliente irá entrar no restaurante?



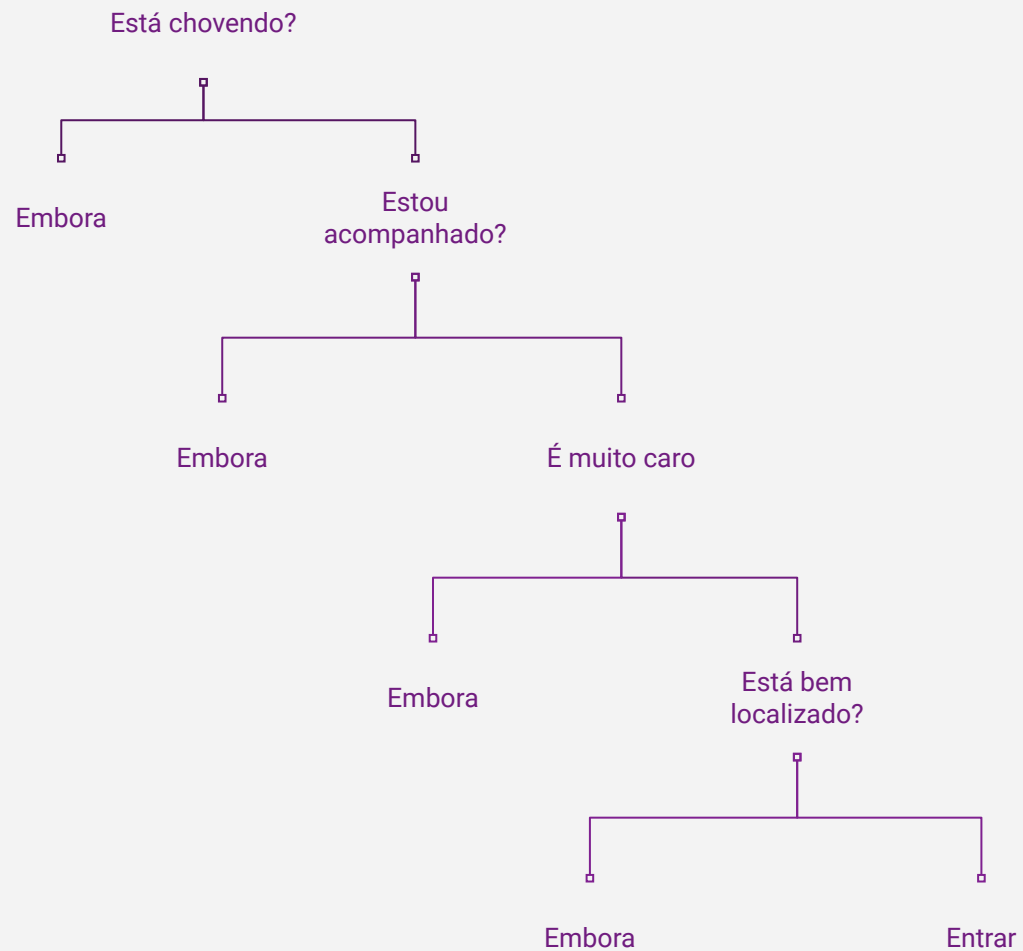
5

Perguntas

Será que o cliente irá entrar no restaurante?

- Está chovendo?
- É muito caro?
- O Restaurante é bom?
- Está bem localizado?
- Estou acompanhado?

Qual é o melhor atributo?

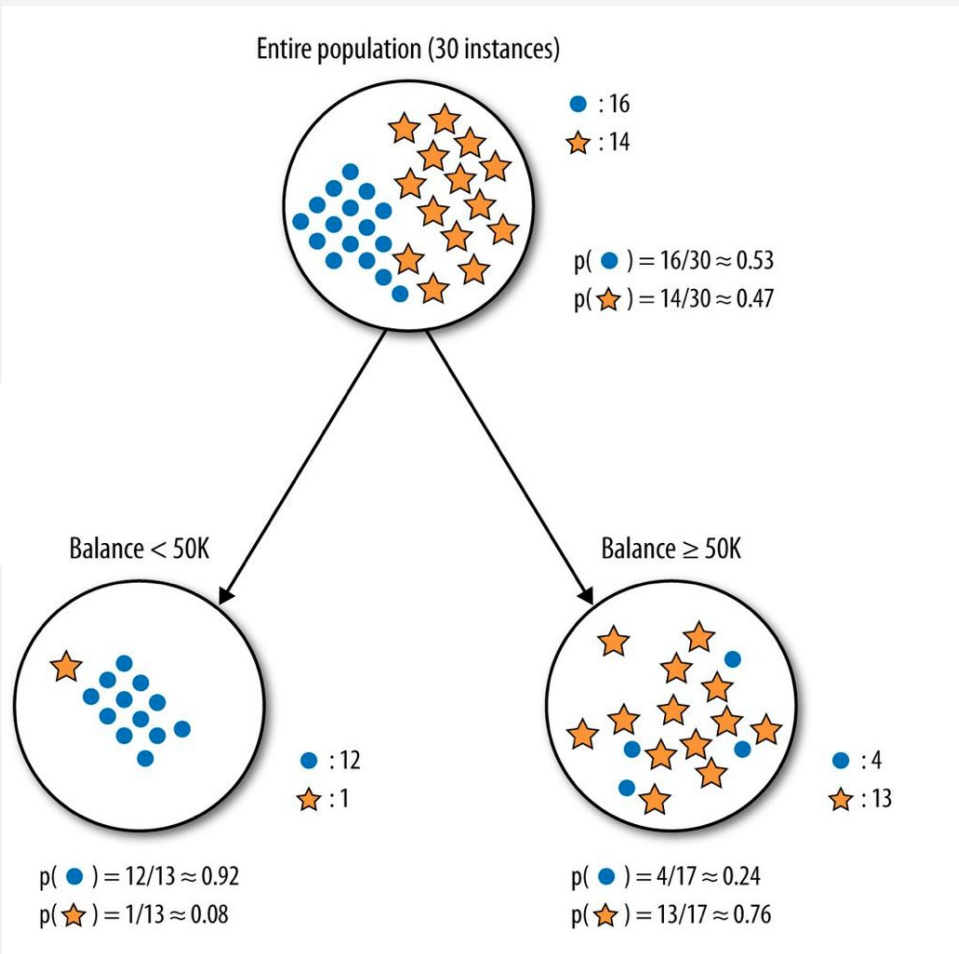


Como podemos quantificar?

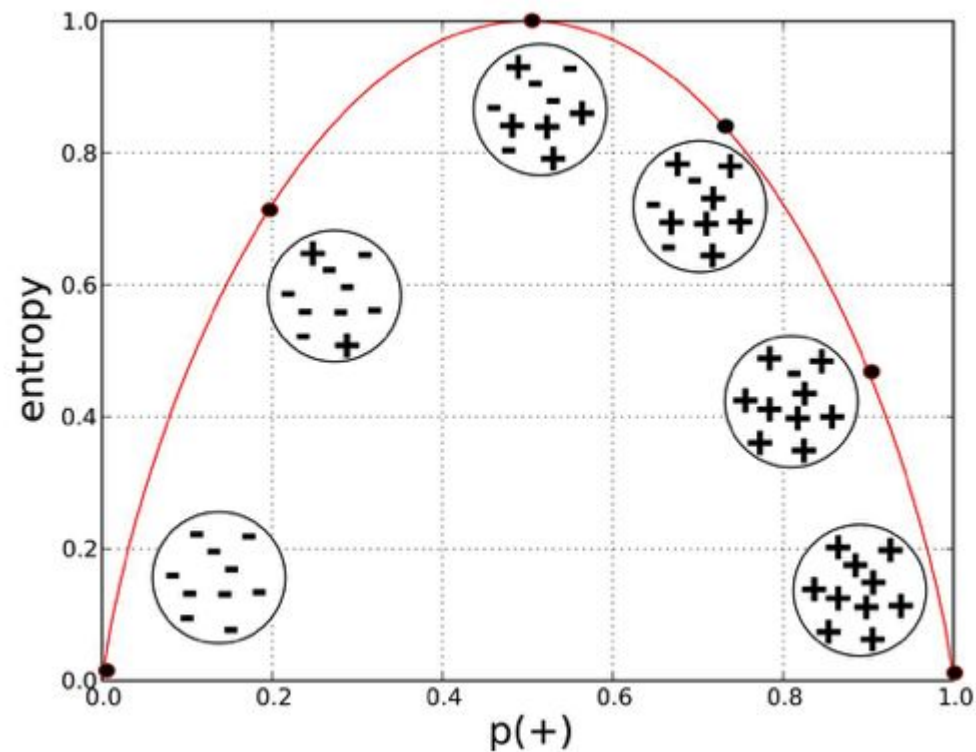
- Entropia
- Gini Index
- Classification Index

Entropy

$$E(S) = \sum_{i=1}^c -p_i \log_2 p_i$$



Entropy



The background features two thin, dark gray lines. One line starts at the bottom left corner and extends diagonally upwards towards the top left. The other line starts at the bottom left corner and extends diagonally upwards towards the top right, intersecting the first line. These lines create a triangular shape on the left side of the frame.

Considerações

Representação e Algoritmo ML

Representação

- Mapeia todos as combinações possíveis
- Pode se tornar infinito
- Aprendizado Especializado

Algoritmo de ML

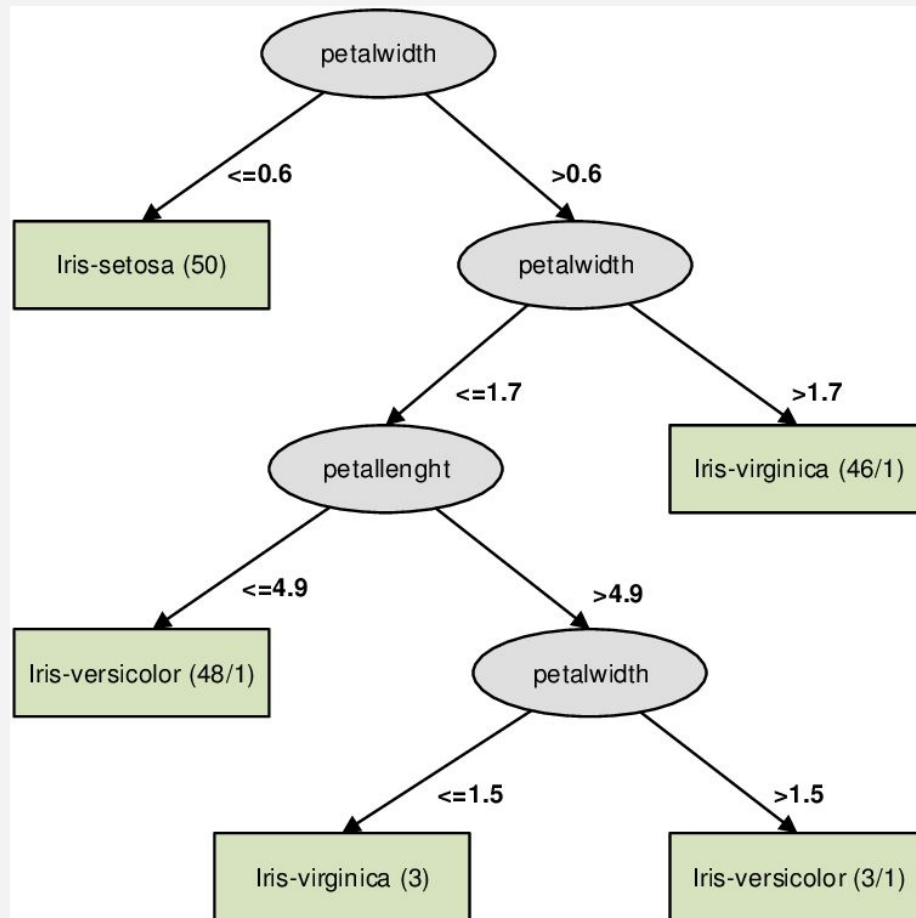
- Mapeia as combinações até o ponto de parada.
- Aprendizado Generalizado

Ponto de Parada (Póda)

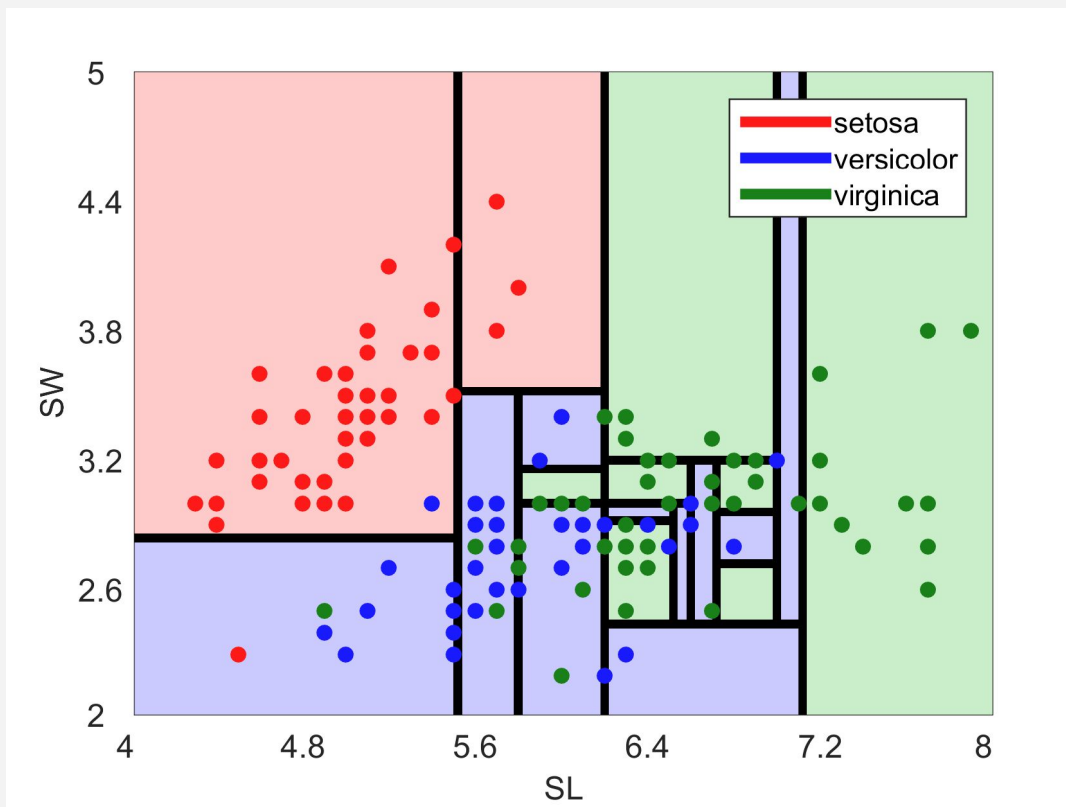
- Quando tudo foi classificado corretamente
- Não tem atributos
- Profundidade Específica

Dados Numéricos

Para trabalhar com dados numéricos nós dividimos todas em faixa de valores.



Dados Numéricos



Vantagens e Desvantagens

Vantagens

- Fácil e intuitiva
- Consegue explicar para especialistas de outras áreas

Desvantagens

- Complexo com muitas variáveis
- Sensível aos dados